Using Polymorphic Microsatellites to Determine the Population Genetics of *Vespula maculifrons*

A Thesis

Presented to

The Academic Faculty

by

Veronica Thompson

In Partial Fulfillment

Of the Requirements for the Degree

BS in Neuroscience with the Research Option in the

School of Biological Sciences

Georgia Institute of Technology

Fall 2019

Table of Contents

Abstract

Social insects have an interesting genetic history and are studied in order to discern how their social behaviors affects their genetic makeup. The eastern yellowjacket *Vespula maculifrons* is one such species whose altruistic behaviors and caste system should negatively affect their genetic diversity but instead has fluorished for many years as a dominant species in their ecosystem. We investigated whether *V. maculifrons* follows the pattern of other social insects in having a small genetic diversity and therefore, a small effective population size. We sequenced seventeen polymorphic microsatellites of *V. maculifrons* of three different years that were chosen in accordance to the temporal method. We performed a Fixation Index test on the data with the three years as subpopulations in order to determine the differences in allele frequency amongst the groups over time. This was done in order to support our theory that *V. maculifrons* has a low amount of genetic diversity, which correlates to low amounts of allele fixation, and therefore a low effective population size. We found that the fixation index was significantly low, which supported this idea that not many alleles have gone to fixation. This would indicate that the effective population size is low because the population is still affected by genetic drift. In the future, a concrete calculation of the effective population size will be performed with combinations of multiple equations that can account for the many unique social traits of *Vespula maculifrons.* This will then help in order to add more information to the gap of knowledge on the fascinating genetic makeup of these unique social organisms.

Introduction

Many social insect populations have been well studied by scientists to measure population dynamics and help them understand the evolution of the species and its sociality. One such social species with an interesting social structure and is unfortunately understudied is *Vespula maculifrons,* the eastern yellowjacket. They are a unique social species and studying their evolutionary genetics is important and can be done through estimating the effective population size in order to look at their genetic diversity.

One way to determine the genetic diversity is by estimating the effective population size. This value is the number of individuals that has the same genetic diversity as the real population which can be a simpler way to track the genetic changes of a population throughout time, without needing to track the entire population, which is often impossible. The effective population size (Ne) is strongly affected by inbreeding and genetic drift and can show how a species evolves through time and how it differs in its genetic diversity. This can be done by measuring the amount of genetic changes at certain microsatellite loci for *V. maculifrons*. If there is an abundance of genetic differences between alleles, it would indicate a small population size that is vulnerable to genetic drift. If most of the alleles have gone to fixation, shown by little to no genetic changes within the alleles of each year, then it would indicate a large population size.

Social species such as the eastern yellow jacket *Vespula maculifrons* are unique in that they have been around for thousands of years and as such have a unique evolutionary history. This project is working towards determining how that evolution of sociality affects the genetic

diversity (2). This will be tested using short tandem repeats in the DNA called microsatellites. The use of microsatellites to determine the effective population size can be applied to other social species in the future. This can also be used to efficiently determine the population size of those insect populations and help create conclusions about the genetic diversity of social species. Prior to this, no other project has used the temporal method to such a large scale to determine the effective population size and as a result, this project will be a good indicator on the success of this method. This can be successfully paired with technology such as genetic markers like microsatellites as well as PCR to amplify and compare DNA of organisms to accurately estimate populations. *Vespula maculifrons* is being studied to determine its effective population size (Ne) over time, which can then be used to determine how social interactions can affect a population through time and help with conservation efforts.

V. maculifrons is a unique social organism and has been studied for not only its social structure, but also because of its evolutionary past. Recently, scientists have been taking an interest in these yellow jackets in order to study superorganismality, a concept that social insect colonies can be thought of as a single higher-level organism (12). This can help show how some altruistic behaviors can be passed on through generations, such as how the colony takes care of only one queen who produces all the colonies' offspring. This queen then partakes in polyandry, of which *Vespula maculifrons*, in particular, does so in high proportion to other social insects, causing higher genetic diversity (2). Also, *Vespula maculifrons* is an example of the organisms that have a trade-off between population size and genetic diversity. Most organisms face a trade-off between lifespan and reproduction ability, but social organisms like *V. maculifrons* have instead evolved to have large population sizes and high reproductive success at the expense of reproductive ability (10). In studying the genetic diversity, one can answer the questions of why this occurs, as well as determine why Lewontin's paradox, which is the phenomenon where the genetic variation of a species does not increase proportionally with the size of a species, is prevalent in these populations (14).

In other studies, *Vespula maculifrons* has been experimented on to determine if the polyandry is beneficial to the society and gives a fitness advantage to the queens (2). Polyandry can also affect how resources are allocated by the population (4), and this should increase the genetic diversity on which these social insects like *Vespula maculifrons* rely. Polyandry also has a unique effect on the yellow jackets because an increase in similarity could be detrimental, as organisms are more likely to support one another if they are more related. With this increase in genetic diversity, the sister workers are less likely to act altruistically, which can negatively affect the entire colony. The effective population size is affected by a variety of different factors and is intriguing to study its evolution throughout history because of its detrimental effects, like a reversal of natural selection (10). The study seeks to investigate this interesting trait and see how it changes throughout time to gain an understanding of why it is still prevalent despite its many consequences.

In this experiment, *V. maculifrons* queens and workers were collected from around Atlanta over a period of 14 years, and then samples from the chosen years 2004, 2006 and 2017 were compared to each other. We ensured that there are multiple generations between each

measured population to account for any confounding variables in accordance with the temporal method (9). We are using microsatellites to track mutations and changes in the alleles over time and use the number of changes in the alleles to first determine if the population size is small or large. In order to have an estimate of the genetic variation and therefore size of the effective population size, we performed a Fst Fixation Index Test with the years as the subpopulations being tested instead of the colonies in order to give us a better estimate of the genetic variation changes over time. We then will determine the effective population size with this data by measuring the mutations in each locus while also accounting for the temporal differences, which has not been done before.

The data from this project can help other researchers, as it can assist them in determining population sizes of other organisms in a method that involves measuring alleles over time. Researchers will then contribute this data to understanding the genetic diversity of social insects. This can help answer evolutionary and genetic history questions such as how to understand genetic diversity of such complex insects, and how this relates to Lewontin's paradox. We will analyze 17 polymorphic DNA microsatellite markers to determine differentiation between alleles. The polymorphic microsatellites can not only determine the effective population size, but can also be studied for how they evolve within a population by measuring how they are either gained or lost among certain individuals within the group. This study, by using microsatellite DNA markers to study the effective population size, can help further investigate on how *V. maculifrons* evolved socially and genetically throughout its population.

Literature review

This study involves determining the underlying effects of traits of sociality that insects like *Vespula maculifrons* have evolved and perfected throughout its life cycle through measuring polymorphic changes and calculation of its effective population size. This study is unique in that no other project has done a genetic analysis of social insects on a large scale by using the temporal method to fully measure the effective population size.

Microsatellite DNA markers are commonly used to determine how genetic diversity has evolved through time with usage of various other insects such as cichlids as shown in Nunney and Elam's population research (9). This method has been used for conservation efforts in various subjects and human studies throughout the years because the DNA markers make up almost three percent of the human genome and are usually conserved during selection (9). Microsatellites have been studied for years because of their unique trait of high rates of mutation, and some can affect phenotypes and can exhibit deleterious traits or traits that are beneficial to the individual. They have been studied mainly in humans, as they could have important medical significance, but studying human microsatellites presents many obstacles because of nucleotide substitution, as explored in Sawaya, *et. al* (2012) in which microsatellites are compared to determine their effect on human evolution and studies how microsatellites are gained or lost throughout time (11).

One example of how microsatellites are commonly studied in the ecological field involves determining the effective population size. The effective population size (Ne) is an important parameter and has been used to study organisms with various sizes of populations because it is an indicator of how populations evolve over time and measures the strength of population's genetic drift (9). Effective population was determined to be a strong indicator when methods for conservation were unsuccessful, so the ratio of genetic evolution to the entire population is often noted as a more useful parameter that would be more effective than simple habitat selection. However, the method was not widely used until popularized by Nei and Tajima's work with *Dacus oleae* (the olive fruit fly) as an alternative to the total population size (N) (7). The effective population size is not as commonly used, especially with insect taxa, because of its complexity in calculation as reported in Watts *et al.*'s research (11). Various common methods exist to determine the effective population size, but comparison studies are lacking in the literature. Their study compared the demographic-based studies with those that used genetic methods and determined that the demographic-based estimate usually overestimated the effective population size. The genetic-based method, in contrast, closely represented the effective population but overestimated the migration rates. Watts *et al.*'s and Nunney and Elam's research suggests that the effective population size is always less than N, and the ratio is often taken in order to determine genetic variation (13, 9). Some issues that have occurred in calculating the effective population (Ne) size in past experiments include that the effective population size relies on generation time and usually equals N/2, with N representing the adult population, as discussed in Nunney's experiment on how multiple mating can affect the effective population size (8).

There have been studies that use this genetic model for determining the effective population size, but there has been an issue with the amount of time between the sampled populations that, when studied, would give an effective and unbiased estimation of the true population size. One such study is by Nunney and Elam that brought into the mainstream a new method called the temporal method (5). The more common approach, especially for long-term genetic studies, was using replicate populations, which were artificially created populations to model the actual population. These were not as successful because they had multiple requirements for its success such that it could only be used in exact replicate populations and did not account for the effects of mutation or selection (5). In contrast, the temporal method is supported by various studies and is shown to account for confounding variables by measuring multiple independent markers and is best for multiple generations, as opposed to the impractical estimations of lethal frequencies which is inherently unsustainable, as shown in Nunney and Elam's research comparing all the known estimating methods at the time (9). This temporal method was first introduced by Nei and Tajima whose formula is more applicable to a wider array of cases for nucleotide substitution of microsatellites, much more than its predecessors (7).

Furthermore, effective population size is a good indicator to help contribute data to working through Lewontin's paradox, which is the phenomenon that genetic variation does not increase linearly with organismal population size (14). This is a well-known paradox within geneticists and *Vespula maculifrons* is a good species to understand more about this because of its social interactions such as its high amounts of polyandry. Polyandry as a practice between organisms is common within social insect structures, especially in *Vespula maculifrons's* family *Hymenoptera*. This behavior has been an important topic to study, as shown in Boomsma's research on how the queens and their male mates are often under differing selection in order to

either minimize relatedness regarding males or maximize their relatedness in the case of the queen due to evolutionary fitness benefits (1).

There have been various studies on polyandry and its effects besides Boomsma, with Goodisman being the main supporter of using *Vespula maculifrons* in its usefulness to determine various social interactions (2,3). These social groups have been studied for many years with members of the social groups cooperating with each other under no apparent fitness gain. Studies have shown that this behavior is likely due to evolution for kin selection, showing the individuals' positive behaviors towards relatives increased to increase their fitness, and most of these studies have focused on the more popular insect, *Apis mellifera*, the honeybee. However, these studies were not as successful since the *Apis* colonies did not produce high amounts of queens, so their reproductive competition does not apply but opened the doors for other studies on more typical social insect colonies, especially *V. maculifrons* (3).

V. maculifrons has a diverse social structure, and they have a known social biology, caste system, and evolution history, so they are the perfect candidate to study. As demonstrated in the research by Goodisman *et. al* (3), many other social insects have multiple families within their caste while *V. maculifrons* is unique in that the colony is led by a single queen that has multiple mates, so the confounding variables of multiple queens is not an issue like in other social insects like *Apis mellifera*. Studies exist on how polyandry may be beneficial to their society and how this behavior may have any effects on the allocation of resources, such as Johnson's study that showed the advantageous effects of genetic diversity on the caste and its resource allocating ability (5). Biologically, *Vespula maculifrons* is a haplodiploid species, where the diploid female workers have a set of chromosomes from both parents while the male drones only contain one set of chromosomes, which makes it interesting to study the effective population size by using patriline analysis. The program MATESOFT has been tested and indicated to be the most accurate and useful method of analyzing paternal relationships, as it accurately infers genotypes of both paternal and maternal and so is now the industry standard (6).

The current study addresses the gap in the research of using microsatellites and the temporal method to determine how the effective population size of *Vespula maculifrons* changes over time due to inter- and intra-colonial effects. This study will allow other studies to do the same with other organisms and will enhance our understanding of the deep social interactions of this unique social organism. We will be using the genetic method using microsatellites to calculate the effective population size and will also be using the temporal method as discussed above to determine its validity in combination with other methods. The other studies have used these methods separately to show their individual validity, but this study will demonstrate how this combined method can fully explain the genetic history of *Vespula maculifrons*. This study and method can then be built upon in the future to inspire other experiments to do the same with other organisms across the world.

Materials and Methods

We intend to extract, amplify, and determine the length of the base pairs of microsatellites of different colonies of the eastern yellowjacket *Vespula maculifrons*. First, over 100 colonies each containing a queen and on average 15 female workers were collected over a period of 10 years around Atlanta according to the temporal method (9) and then frozen to have their DNA extracted. The extraction was performed using the omega bio-tek E.Z.N.A. Tissue

DNA Kit on multiple colonies by using the thorax of the individual as a tissue sample. We then create simplex and multiplex Polymerase Chain Reactions using a mix of one or more of 17 polymorphic microsatellites markers to amplify the desired locus. These markers are RUFA5, RUFA12, RUFA19, VMA3, VMA6, VMA8, LIST2002, LIST2003, LIST2004, LIST2007, LIST2008, LIST2010, LIST2013, LIST2015, LIST2017, LIST2019, and LIST2020. They were found previously in Dr. Goodisman's research and are conserved in certain rates throughout the population and so can be used to study the genetic changes succinctly instead of using the entire genome. The PCR mixture for a 1:1 simplex recipe is 14 µl extracted DNA, 6.90 µl PCR water, 1.50 µl 10X PCR Buffer, 2.40 µl MgCl2, 1.20 dNTPs, 0.75 µl Forward Primer, 0.75 µl Reverse Primer, and 0.50 µl Taq polymerase. When performing a multiplex, another 0.75 µl Forward and Reverse Primer are added, and the total amount added is subtracted from the PCR water to equal 14 µl. These are common ingredients in successful PCR reactions and this recipe has been found to produce the clearest results when amplifying the DNA.

This PCR was run to multiply and amplify these microsatellites to make millions of copies that will then be sequenced. This PCR product was then run using gel electrophoresis to determine if the amplification was successful and checked to determine if the correct base pairs appear. The resulting successful PCR fragments were subsequently run on an ABI 3100 Genetic Analyzer. This tells us the specific base pairs of each individual and if the individual is homozygous or heterozygous, which will help us with further patriline analysis where we determine the paternal parents of all the female workers. Alleles were scored using GeneMarker v 4 and then we used the coding program GENEPOP v 4.3 for the intercolonial tests. An Fst Fixation Index test was performed using the years as the subpopulations and all of the years total as the total population. Doing this will give us an estimate of the genetic variation over time which will then give us a proper estimate of whether the effective population size is small or large and will then influence our future studies. We also performed a patrilinity test which helped us determine the range of inbreeding within colonies and will be finished once all of our data is collected. The patrilinity helps determine the relativity of each of the generations and will contribute to determining the mutation rate. This data will then be used to determine the effective population size with a combination of multiple equations and the collected data.

Results

DNA was chosen and extracted successfully from four colonies collected in the year 2006, ten colonies collected in the year 2008, and twenty colonies from the year 2017. The seventeen loci were sequenced for the colonies 3, 7, 8, and 9 from 2006. Colonies 63, 65, 66, 67, 68, 72, 74, 76, 77, and 78 were sequenced for the year 2008 on the same loci. Also, colonies 193, 195, 196, 198, 199, 200, 201, 203, 204, 205, 206, and 221 were sequenced for the year 2017 on the chosen loci. Each of the years was used as a subpopulation when calculating the fixation index and then they were compared to the sum of the data from all three years as the total population. The data determined that there was largest Fst value, in Rufa5 and the lowest value in LIST8. The total Fixation index for the entire data was 0.1293 (Table 2). A patriline parentage analysis was then performed for two of the colonies, with 17 loci successful for colony 193 and 15 out of the 17 loci successful for colony 193.

Figures

Ind	RUFA5	RUFA12	RUFA19	VMA3	VMA6	VMA8	LIST2002	LIST2003	LIST2004
Q1	139,139	82,84	207,207	272,266	286,306	250,250	171,191	205,201	159,159
M1	139	82	209	268	294	242	171	200	150
M2	139	82	207	276	274	234	177	X	153
M3	141	88	199	268	296	258	178	204	153
M4	141	88	199	268	296	258	178	204	153
M5	139	82	207	274	274	234	177	204	155
M6	139	82	207	274	274	234	177	204	155
M7	139	82	207	274	274	234	177	204	155
M8	139	82	207	274	274	234	177	204	155
M9	141	88	199	268	296	258	178	X	153

Ind	LIST2007	LIST2008	LIST2010	LIST2013	LIST2015	LIST2017	LIST2019	LIST2020
Q1	155,153	147,151	193,189	206,207	166,166	164,166	139,135	234,238
M1	154	147	185	204	167	158	139	246
M2	151	141	189	204	166	164	150	236
M3	155	153	191	207	168	166	139	234
M4	155	153	191	207	168	166	135	234
M5	151	141	193	204	166	164	150	236
M6	151	141	193	204	166	166	150	236
M7	151	141	189	204	166	164	150	236
M8	151	141	189	204	166	166	150	236
M9	155	153	189	206	168	164	139	234

Table 1. Male Generated Alleles Based off Queen Alleles from Colony 198: This table lists the alleles at each locus (17 total) for possible male mates of the queen (alleles also displayed). It was generated by MateSoft® FQ Patriline Analysis. The primer names for each locus are listed in bold starting with R5 and ending with L20. The allelic information for the queen displays the two alleles of the queen for a locus (amplified by the primer indicated) in terms of base-pair length. The information for the allele donated by the male to the offspring is listed in terms of base pair length. The base pair lengths at VMA6 for the queen was missing, so a MateSoft® FQ Patriline Analysis was performed to determine the alleles of the queen at this locus.

Locus	Fst value
RUFA5	0.6339
RUFA12	0.1406
RUFA19	0.1189
VMAC3	0.0692
VMAC6	0.0625

VMAC8	0.0734
LIST2	0.0625
LIST3	0.1604
LIST4	0.0381
LIST7	0.0377
LIST8	0.0049
LIST10	0.2221
LIST13	0.0178
LIST15	0.1702
LIST17	0.1719
LIST19	0.2121
LIST20	0.1154
Total	**0.1293**

Table 2. Fixation index values of each of the years for each microsatellite. his table lists the Fixation Index values for each polymorphic microsatellite locus tested on all the collected colonies from 2006, 2008 and 2017. The primer names for each locus is listed on the left while its correlating Fst value is listed on the right column. All the fixation index values were positive. They were calculated using the coding program GenePop v 4.3. Also, the total Fst value is shown above in bold to be 0.1293.

Discussion

According to the results of patriline analysis of colonies 193 and 198, the patrilines were successfully calculated using the MateSoft program (Table 1). This will assist in the future of the project to calculate the relationships among the colonies to better compare the differences in the allelic mutations once more data is received. This information will also help us track the number of males that mate with the queen and determine the haplotype of the offspring for future analysis.

According to the data, the fixation index total value is 0.1293. This is a very low genetic variation and indicates a high amount of shared genetic material, and very little fixed alleles. The microsatellite RUFA5 had the highest fixation index with a value of 0.6339, indicating more fixation of the allele, and with our data, LIST8 has the smallest fixation index value of 0.0049 (Table 2). This information shows that there is fluctuation in the genetic diversity among the different microsatellite alleles across the years and there is little fixation in the population. This indicates that overall, the effective population size is moderately small and affected by genetic drift. Also, the Fst value supports that our theory that the effective population size will be a smaller number, as supported by Romiguier's work (10). The results from the patrilinity test will be used to guide the progress of this experiment in order to analyze the data by helping determine the haplotype and identity of the male mates. In order to fully calculate the effective population size of a unique organism like *Vespula maculifrons*, we will need to combine multiple equations and theories from other experiments and a weighted equation that accounts for the polyandrous nature of the breeding females. In conclusion our results confirm our original idea

that *V. maculifrons* will have low genetic diversity and fixation which supports our hypothesis that this organism will have a small effective population size. This will then contribute data to better solving Lewontin's paradox and adding more data to fill the gap of information about the population genetics of *Vespula maculifrons.*

References

1. Boomsma JJ. 1996. Split sex ratios and queen-male conflict over sperm allocation. Proceedings of the Royal Society of London. Series B: Biological Sciences 263:697-704.
2. Goodisman MA, Kovacs JL, Hoffman EA. 2007. The significance of multiple mating in the social wasp Vespula maculifrons. Evolution 61:2260-2267.
3. Goodisman MA, Kovacs JL, Hoffman EA. 2007a. Lack of conflict during queen production in the social wasp Vespula maculifrons. Molecular ecology 16:2589 -2595.
4. Johnson EL, Cunningham TW, Marriner SM, Kovacs JL, Hunt BG, Bhakta DB, Goodisman MA. 2009. Resource allocation in a social wasp: effects of breeding system and life cycle
5. Johnson MS. 1988. Founder effects and geographic variation in the land snail Theba pisana. Heredity 61:133.
6. Moilanen A, Sundstroem L, Pedersen JS. 2004. MATESOFT: a program for deducing parental genotypes and estimating mating system statistics in haplodiploid species. Molecular ecology notes 4:795-797.
7. Nei M, Tajima F. 1981. Genetic drift and estimation of effective population size. Genetics 98:625-640.
8. Nunney L. 1993. The influence of mating system and overlapping generations on effective population size. Evolution 47:1329-1341
9. Nunney L, Elam DR. 1994. Estimating the effective population size of conserved populations. Conservation Biology 8:175-184.
10. Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, Ravel S, Ballenghien M, Cahais V, Bernard A, Loire E. 2014. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. Journal of evolutionary biology 27:593-603.
11. Sawaya SM, Lennon D, Buschiazzo E, Gemmell N, Minin VN. 2012. Measuring microsatellite conservation in mammalian evolution with a phylogenetic birth–death model. Genome biology and evolution 4:636-647.
12. Taylor D, Bentley MA, Sumner S. 2018. Social wasps as models to study the major evolutionary transition to superorganismality. Current opinion in insect science.
13. Watts PC, Saccheri IJ, Kemp SJ, Thompson DJ. 2007. Effective population sizes and migration rates in fragmented populations of an endangered insect (Coenagrion mercuriale: Odonata). Journal of Animal Ecology 76:790-800.

14. Ellegren H, Galtier N. 2016. Determinants of genetic diversity. Nature Reviews Genetics 17:422.